

## Haile Quals day 4

*Some of your readings cover the development of expertise (or skill) from a cognitive perspective (mind rather than brain)*

- *According to these readings, what are differences between experts and novices (those who are highly skilled and those in the initial stages of learning)?*
- *What changes as a person becomes more skilled or expert in a particular domain?*
- *Are there differences/disputes/disagreements with respect to what is taking place in the development of expertise (or skill) among the views represented in your readings? If so, explain them and offer your own opinion on the topic.*
- *Finally, what are the implications of this issue for our understanding of human cognition in general?*

Most skill learning theories bear the same hallmarks of expert performance. Experts can execute tasks efficiently with little mental effort and at much higher levels of accuracy than novices. In most theories, this is a multi-stage process but the number of stages and the reason for the improvement tend to vary or are non-specific. Current theories in the cognitive sciences seem to have been influenced by the much earlier Fitts and Posner (1967), Schneider and Shiffrin (1977), and Anderson, (1982) models so these models are discussed in this essay. However, there are nuanced but vital observations about skilled performance that do not appear in the cognitive and computational models of skill learning. For instance, Kahneman and Klien (2009) recognize that expertise or highly developed skills are domain specific and ill equipped to meet dynamic environments or those with a large number or probabilistic cues. Additionally, access to memory and information representation seem to also change in a domain specific manner with the development of skill (Eriksson and Kintsch, 1995). This throws into question the completeness of the models of skill learning discussed here since most of them were developed in tightly controlled lab environments and are limited in their predictions about performance in complex, naturalistic situations. These observations require that we take a broader view of what the purpose of skill learning is and the tug-of-war between efficiency in performance and flexibility in dynamic environments.

There are two big questions in skill learning: what does expert performance look like compared to novice performance? And how does this expert performance develop? These two questions are intimately linked because it seems like some of the models of skill development have started by observing skilled performance and tried to engineer a cognitive and psychological explanation for how they came about.

There are many terms in psychology that describe a minimum number of 2 psychological states concerned with task performance and these seem to rely on the features of different stages of skill performance. Expert performance is usually described as being fast, sparing of cognitive

resources, effortless, virtually error free, automatic, and relatively inflexible. This performance is almost habitual. Novice learners are often characterized by their effortful, slow, and error-prone performance. This pattern of behavior is largely because novices have to pay attention to all aspects of the environment and their response patterns (Fitts and Posner, 1967; Schneider and Shiffrin, 1977; Anderson, 1982; Tenison et al., 2016).

Two varying schools of thought on expertise reveal nuances about the skilled performance described above that build a better, more complete image of expert performance. The first of these two, the Naturalistic Decision-making group of thinkers study successful intuitive judgements made by experts in naturalistic and complex environments. The second, heuristics and biases group of thinkers, highlight that those intuitive judgements, even by experts, can be wrong and a formal computational strategy should be preferred in complex situations. Both groups agree that stable and predictable environments are well-suited for development of skill and, performers make more correct intuitive judgements as they gain more experience in such environments (Kahneman and Klein, 2009). However, they seem to diverge in ideas when the environment is complex. Here complex means, the environment has more cues than the experts are aware of, some of these cues may be probabilistic and some cues maybe valid only sometimes. In these situations, experts might rely on intuitions developed through exposure with similar situations but have a high chance of producing the wrong response. The Naturalistic decision-making group aims to find the cues that led to successful, implicit, intuitive judgements while the heuristics and biases groups recommends a more effortful and formal reasoning and decision-making process that considers more global probabilities. For instance, a clinician might be able to make more accurate diagnoses of difficult cases by performing a statistical analysis that takes into account the probability of occurrence of a specific ailment given the patients' demographics. This is opposed to intuitively deciding on a diagnosis using only the symptoms the patient is displaying and the clinicians unique set of experiences, which may not always be reliable (Kahneman and Klein, 2009).

The best response to a situation is one that balances efficiency with accuracy. There is not a perfect strategy for how to respond in complex situations. Skills are necessary for fast and efficient response to predictable situations. But these fast responses come at the expense of accuracy if the environment ceases to be predictable or is unreliable some of the time. There are not many formal models for learning in unpredictable environments, and the popular models described here are largely built for stable environments. In the case where the environment isn't reliable, the usual assumption is that learning does not occur, and a laborious solving of the problem must be performed as we will see with the Schneider and Shiffrin (1977) studies.

With proponents of 2-system theories, early-stage performance is termed *controlled processing* and late-stage, or expert performance is known as *automatic processing* (Schneider and Shiffrin, 1977; Hill and Schneider, 2006). These theories are influenced by early visual attention and visual search experiments. Schneider and Shiffrin's 2-system theory was demonstrated by visual search tasks where subjects were required to find and report items from a specific category and ignore other categories of objects. In varied mapping experiments, where

the target category changed trial-by-trial, subjects had to rely on serial search methods that required attentional engagement and there was not much improvement in speed (controlled processing). But there was an improvement in speed and accuracy with subjects who experienced consistently mapped visual search experiments, where the target category does not change from one trial to the other, because they were able to engage in automatic processing. Therefore, controlled processing must be engaged in new and constantly changing situations (Hill and Schneider, 2006). Additionally, controlled processes are affected by task load and fatigue whereas automatic processing is largely not, pointing to differential use of available cognitive resources.

Schneider and Shiffrin's controlled processing is similar to what Kahneman (Kahneman and Klein, 2009) calls *System-2*, which is a slow, effortful problem-solving system which will have better success in decision making in complex situations at the expense of speed of making a decision. To reiterate, controlled processes rely on domain-general attentional and executive processing because the subject has to deploy attention to search stimuli in the environment and in long-term memory. But once sufficient training occurs in a reliable environment, this top-down, and effortful control and attention network is not needed (called *System-1* by Kahneman). Automatic processes are represented in a non-specific long-term memory store where the representations (nodes, according to Schneider and Shiffrin) are activated automatically by current stimuli and responses are deployed without the subject directing them (Schneider and Shiffrin, 1977). In this case, the above-mentioned domain-general attentional resources are free to be utilized elsewhere. This notion has been exploited by various attention and skill learning studies using dual-task experiments to demonstrate that skilled performance does not rely on top-down attention (e.g., comparing expert golfers with novices while they memorized items from a list delivered aurally; expert golfers recalled more words from the list compared to novice golfers - Beilock et al. 2002). This controlled and automatic processing has gone on to influence multiple lines of research, including more recent attempts to find more evidence for how this is implemented in the human brain through network neuroscience (see Hill and Schneider, 2006 for a review).

However, the 2-system theory described above does not well characterize how skill transitions from requiring a high level of control to automaticity, and the systems described seem non-specific. Other concurrent theories describe a multi-stage development of skill that are more specific in the types of long-term memory that guide novice and expert performance. The most notable of these are theories by Fitts and Posner (1967) and Anderson (1982). Both models are influenced by computation and computer structure theories typical to their time from the 1960s to the 1980s, that have prevailed in modern theories of cognition.

The Fitts and Posner (F&P) (1967) model of skill acquisition has broad strokes that are like the Schneider and Shiffrin model but goes a little deeper with a hierarchically organized system that contains 3 learning stages: 1) Cognitive Phase, 2) Associative phase, and, finally, 3) Autonomous phase. The F&P model also tracks the familiar progression from effortful cognitive engagement in the early stages of learning to the more effortless and automatic late-stage

learning. It should also be noted that progression through the stages occurs gradually and not in a staccato fashion even though they have forwarded distinct phases.

This model conceptualizes skill development as a streamlining process that constructs response ‘routines’ that are made-up of smaller, relevant, and flexible ‘sub-routines’ or habits. It may be helpful to think of subroutines as the learner’s store of stimulus-response maps and strategies and, they are like Schneider and Shiffrin’s long-term memory nodes. For instance, learning to write functions in python might be built from typing skills, arithmetic skills and several other, more general, problem-solving and organizational ‘subroutines’ or habits. But the F&P model includes an additional hierarchical structure here that also resemble computing terminology, sub-routines. Sub routines are pieces of code or functions that perform repeated tasks and make up larger programs. Similarly in human cognition, F&P (1967) argue that the first cognitive phase is learning which subroutines are relevant, out of a multitude of available subroutines. This phase requires attending to and representing relevant features of stimuli, specific cues, and responses, which requires effort. During this stage, the learner normally applies inappropriate subroutines which lead to errors and poor performance probably because the learner lacks the knowledge of which subroutines are optimal, and some other new ones need to be generated. There is a reference to the learner’s memory, and executive and attentional resources but not much else is described along the more influential descriptions of declarative and procedural memory/learning processes that come later in other models.

During the second stage, the associative phase, new subroutines are created, and new routines that were constructed from older subroutines are tested by evaluating their success rate with respect to feedback from the environment. Then a process of elimination is performed where inappropriate subroutines that result in errors are tuned out or eliminated (this process of *tuning* will also appear in Anderson’s skill learning theory, stay tuned). Successful routines are practiced and the number of activated inappropriate subroutines are reduced along with an appreciable reduction in errors.

The final autonomous phase occurs once the new routines are established and become automatic. There are virtually no mis-applied subroutines, and the learner no longer needs to deploy attentional and executive decision-making resources to parse available routines and selects only those that led to success. In other words, the learner is able to perform the learned skill almost by instinct, like how experienced drivers don’t have to effortfully monitor street signs and other vehicles on the road to successfully navigate. There is virtually no explanation of how this transition through the learning stages occur in terms of human learning and memory mechanisms that I could find in the early text. For instance, are all subroutines that support new skills procedural long-term memory? There is, however, more explanation of behavioral features of decreases in errors and increases in speed and efficiency using this framework. Anderson (1982) goes a little further to establish a more detailed progression of skill learning that further develops in more recent times.

Anderson’s model of skill learning, while it shares some features with its early contemporaries, describes automatization of skill more specifically. It also forms the basis for the

expansive, popular, and influential cognitive architecture, Adaptive Control of Thought-Rational (ACT-R) (Ye et al., 2018).

In Anderson's model (1982), all skills start out as declarative information also known as factual knowledge. Declarative instructions or rules stored in long-term memory, or newly acquired from the environment (e.g., reading instructions for a specific recipe from a cookbook), are attended to and the instructions are carried out laboriously. But factual knowledge in declarative memory only indirectly affects behavior. An additional procedural system interprets the instructions and carries them out. This procedural system contains simpler, domain-general *procedures* for carrying out common tasks, not unlike the Fitts and Posner (1967) subroutines. This is a process that is resource intensive and needs engagement with attention and verbal rehearsal of instructions that are effortfully retrieved from long-term memory. At least at this early stage in skill learning, Anderson describes this as requiring space in working memory, a concept not discussed in much detail by both the Schneider and Shiffrin and, Fitts and Posner models. This early stage also aligns with observations of novice performance of slow, error prone work. For instance, with our cookbook example, the learner can read instructions, maintain them in working memory and carry them out. This will also require effortful monitoring of the outcomes of the steps. The learner may also attempt to memorize all the steps and outcomes and access them from memory when needed. Here, it is clear that performance will be slow and not without errors.

With practice, through a process of *knowledge compilation*, the learner will become less reliant on direct recall of instructions from declarative memory and laborious interpretation by the procedural system via working memory (Anderson, 1982). Compilation has two distinct sub-processes. The first, *composition*, describes the process of combining sequences of procedures into fewer procedures. For example, if the cookbook instructions call for measuring a series of ingredients and combining them in a specific order all these actions will be represented by individual procedures. After practice, these procedures might be combined into 1 or 2 procedures that might just hold the instructions "measure ingredients" and "mix ingredients". These new procedures will require retrieving details from LTM. The second sub-process, *proceduralization*, rebuilds these combined procedures in such a way that they no longer require retrieval of information from long-term memory – they simply become sets of efficient stimulus-action maps. This is very similar to F&P (1967) association phase and results in speed-up of performance and reduction in use of working memory resources.

Anderson's skill learning theory has evolved over the years to seem more like human cognition than a computing framework. Knowledge compilation, in my opinion sounds very much like early computing concepts and terminology, though it goes a step further into explaining what occurs during skill development. This representation of two systems, the first that contains declarative knowledge and the second that learns *how* to act on that knowledge and eventually become independent of the first, much resembles how human cognition is organized. This is so at least from the point of view of numerous studies with amnesic patients. These studies have shown that damage to certain brain regions resulted in loss of some but not all

memory functions. For example, patients with amnesia, can learn new procedural skills but not new declarative facts, or remember that they learned the skill (e.g., Packard and Knowlton, 2002; Anderson, 2002).

Anderson's cognitive architecture, ACT-R, has evolved through the years and was much informed by human and animal neuroscience studies (Anderson, 2007). ACT-R is modular and contains modules for sensory perception, attention, declarative memory, and procedural learning. But it also includes a system of goal representation. But I will discuss only the aspects of ACT-R that are relevant for skill learning, mainly in the procedural system.

Anderson's procedural system (Anderson, 1982; Anderson, 2002) performs the bulk of the "learning" in skill learning. The functions of the procedural system largely explain how skills develop and their behavioral hallmarks.

Skilled performance resulting from practice is a gradual and slow process. It cannot be explained by the rapid learning that we would see in declarative systems. Therefore, this must occur slowly and with reward through reinforcement learning (Anderson, 2002). Reinforcement 'teaches' the learning system which procedures are useful. If a series of procedures are utilized to interpret declarative information and produce a behavior, those procedures are rewarded. This increases their chances of being utilized again given the requirements from the environment (Anderson, 2002). This implementation of reinforcement learning in the procedural system goes a long way to bridging the gap between the model and a plausible biological implementation of learning, much further than the other discussed models.

Recent studies have used the ACT-R cognitive architecture to examine skill development in further detail. These studies have also overlaid ACT-R onto Fitts and Posner's (1967) three phases (cognitive, associative, and autonomous phase), resulting in improved characterization of where improvement occurs. For instance, Tenison et al. (2016) further divide the F&P phases into three stages: Encoding, Solving, and Responding. They used an experiment where learners were given a specific mathematical problem with instructions on how to solve it. During the encoding stage, learners represent the problem in working memory. In the solving stage, learners retrieve the rules for solving the problem from declarative memory and apply it. And finally, they simply report the answer by typing out the computed numbers. Tenison et al. (2016) argue that the amount of time a learner spends in the encode, solve, and response stages change depending on which F&P phase — cognitive, associative, or autonomous — the learners are in. As learners practiced, they spent less time in the 'solve' stage because answers to the problems may be readily retrievable, and so, spend most of their time encoding and responding. Once this happens, they are more likely to be in associative phase than either of the cognitive or autonomous phases. Proceduralization can be so streamlined, with more practice, that retrieval of a response will not even be necessary as the response will be represented by motor patterns (Tenison et al., 2016; Anderson, 1982). This very rapid encoding of the problem and rapid, effortless response means that the subjects have reached an autonomous phase. Tenison et al., (2016), demonstrate this transition using behavioral, cognitive modeling and neuroimaging techniques.

However, we often surmise that a transformation in the representation of domain knowledge (from declarative representation to procedural, for example) only, leads to speed-up in performance, perhaps due to a subsequent disengagement of some cognitive mechanisms like working memory and attention as in the previous models. But most complex skills require maintenance and manipulation of new information that have not likely been automatized. For example, we can learn to expertly perform the long-division algorithm to divide a pair of numbers but the numbers to be divided often change, which requires maintenance and manipulation of novel items. This is evident even in the Tenison et al. (2016) study above where a plateau in performance speed of the pyramid problems is observed after the subjects master that specific algorithm but must process new numbers. This presents a small gap in our understanding of skill development. Furthermore, Ericsson and Kintsch (1995) argue that cognitive mechanisms like working memory, that are normally capacity limited must be augmented or made efficient to handle relevant information, as needed, through development of expertise. Efficiency here could mean that experts have better filtering ability of irrelevant information. They further argue that this change is domain specific, mainly because storage in and rapid access from LTM during performance requires building domain-specific schemas. It must be stressed that such augments are not global or domain general. To build on the previous example, as we gain more expertise with long-division we might develop very efficient methods for retrieving values from the multiplication table and hold many more values in working memory as we compute the steps.

The differences here are slight but maybe vital for a more complete image of skilled performance. Elimination of retrieval from long-term memory or reduced use of working memory might not be the final word on skilled performance as most of these other models suggest. But it is also likely that changes in how these normally capacity limited cognitive functions have upper limits - processing new items, even while using a highly skilled algorithms is not fully automatic and some effortful maintenance and processing is needed, which might explain the speed plateaus reached by Tenison et al.'s subjects.

In summary, skill learning is a multi-stage, multi-component refining of responses or problem solving that is characterized by increase in accuracy and speed, associated with significant decrease in errors and cognitive effort. The discussed models and many others characterize how this occurs by attempting to explain how these measurable behavioral changes might occur. But a lot of them seem to be limited to highly predictable environments where learning certainly happens. A better picture of skill learning should evaluate how learning and expert decision making occurs in complex environments. It is clear that meta-learning changes occur – i.e., storage and retrieval of information seems to be improved in inflexible ways – but the focus of most models have been in transformation of information processing only during skill learning. Lastly, it seems like a lot of the early models were very influenced by computer architectures and computing concepts. They have forever altered how we think about cognition. We are likely to use computing terms like retrieval, storage, buffers, information processing, and

compiling, mainly because of this legacy but they have allowed us to think more concretely about cognition as well.